
**Virtual Institute of Microbial Stress and Survival**
**DOE Genomes To Life Project**
**Progress Report: September, 2003**

## I. Overview

The objective of this monthly progress report is to provide an update of the technical and administrative actions from the previous month as well as forecast upcoming progress for the VIMSS Genomes to Life Project. I want to remind everyone how important to make sure everyone is communicating. The discussion boards (http://genomics.lbl.gov/~aparkin/discus) provide a forum for people to ask questions about direction of the project, priorities, and technical issues that can be read and answered by the entire group. I know email is often the most efficient means but it does privatize some of the important communications. Also, posting project data and information to BioFiles (https://tayma.lbl.gov/perl/biofiles ) is EXTREMELY important. We are in the process of adding user help files to BioFiles – if you have user questions, please contact Keith Keller (tel: 510.495.2766 or email: kkeller@lbl.gov). This is the best metric I can give to the DOE leadership that we are making progress aside from the VIMSS website. Please make us and yourselves visible by donating data and information to the website.

## II. Applied Environmental Microbiology Core

*LBNL*

**SR-FTIR.** We continue to modify the existing SR-FTIR spectromicroscopy apparatus to study *Desulfovibrio vulgaris* under anaerobic conditions. Continued to develop protocols for producing *Desulfovibrio vulgaris* biofilm on reference surface and on different types of mineral surfaces. Continued to establish the IR spectral baseline of *D. vulgaris* biofilm under anaerobic conditions. We have also been using collaborating with Carolyn Larabell to examine the stressed *D. vulgaris* using x-ray tomography. This technique looks very promising and we are repeating experiments now for verification of the sheath formation during oxygen stress.

**Biomass Production.** The first biomass deliveries were made this month from the $O_2$ stress experiments. Details of the experimental protocol and discussions of the problems and success have been posted on Biofiles. Four separate experiments were run in order to deliver sufficient cells to all groups. Three of these experiments were successfully completed, the fourth aborted due to an air leak causing the resazurin in the medium to turn pink during the lag phase. Plate streaking (LS4D and TSA) and microscopic investigation of the cultures did not reveal any contamination problems.

Experiment 1: cells delivered to UCB, Sandia
Experiment 2: aborted
Experiment 3: cells to Diversa
Experiment 4: time course samples to ORNL

The frequency of the experiments is exacerbated by the lengthy lag phase period. Primary cultures are inoculated from freezer stock into freshly reduced LS4D medium. These cultures grow to an OD of 0.4 after 3 days. This culture is then used to inoculate the experimental bottles. The inoculated bottles are incubated for 24 hours, or until the OD reached 0.3, after which the bottles are sparged with either air or N2 for five hours.

Several issues and lessons learned came from the September biomass production, and it is anticipated that more will be discovered as sample processing is completed. At LBNL, one fundamental problem discovered is the accurate reading of OD with the presence of titanium citrate reductant, iron sulfides, and cell mass, with the problems mostly stemming from a lack of an appropriate blank sample. This was overcome for most measurements by using a filtered sample as the blank. The protein assays are also proving to lack sensitivity, especially during lag phase when biomass is low. Alternative assays are being evaluated which have a lower working range.

The explanation of the lengthy lag phase will be investigated in the coming weeks. Since this lag phase was not observed when using Baars medium, differences between the Baars and LS4D will be targeted. Possibilities include the components present in yeast extract, the vitamins, or the titanium citrate reductant.

Other progress includes the receipt of both FairMenTec bioreactors. However, we are still waiting for the probes and software, and an additional pump, as well as installation of the system.

***University of Washington***

Previously established syntrophic co-cultures of *Desulfovibrio* strains (Hildenborough and PT2) with *Methanococcus maripaludis* were transferred into fresh McCm medium amended with 20mM of lactate and 80%/20% nitrogen carbon dioxide mixture in the head space. Co-cultures were incubated in the dark at $30^0$C without shaking and monitored over time for optical density ($OD_{600}$), $CO_2$, and methane accumulation. All co-cultures grew and produced methane (Fig. 1.). The maximum growth rates for both co-cultures were comparable (50 hr doubling time within 100 and 200 hours of growth). Anions and organic acids in the medium were analyzed by ion chromatography using a Dionex 500 system equipped with AS11HC column. Less than 2mM of lactate and 16mM of acetate were detected in the medium after 190hr of growth of *D. vulgaris* and *M. maripaludis* in co-culture. 7mM lactate and 13mM of acetate were detected in the PT2 and methanogen co-culture medium. This is in reasonable agreement with the formation of one acetate molecule per lactate molecule. To determine if the *Desulfovibrio-M. maripaludis* co-cultures were limited by lactate, additional lactate was added to both co-cultures after flushing the headspace with fresh 80%/20% nitrogen carbon dioxide mixture. Methane accumulation, optical density and metabolite concentration was monitored over two hundred hours. Methane accumulation was observed in the headspace of both *Desulfovibrio-M. maripaludis* co-cultures amended with lactate but not in a minus-lactate control. Both accumulated methane at comparable rates, indicating that lactate was not limiting.

The fixed bed bioreactor system was inoculated with a pure culture of *D. vulgaris* Hildenborough and the process of biofilm formation monitored for several weeks. A dark biofilm accumulation became visible on the surface of the glass beads by the end of the first week, loosely attached to the bead surface and very easy to dissociate. Several portions of glass beads were removed after 3 weeks of growth. At this time the reactor was consuming approximately 8mM of 16mM of lactate in the feed. The biofilm was dissociated from the beads by vortexing and recovered cells quantified by: 1) direct counting using a Petroff–Housser chamber and 2) dry weight by collecting cells on a Millipore 0.22 μm filter. These measurements suggested that approximately $3x10^6$ cells had colonized each bead.

We have continued the isolation of sulfate reducing strains from FRC area 2 enrichments. Single colonies grown in agar amended with lactate, acetate or hydrogen and carbon dioxide were transferred to fresh agar using the same growth substrates.

TRLFP analysis of 16S rRNA genes amplified from sulfate reducing enrichments growing on different substrates revealed a single dominant fragment for enrichments grown on either acetate or on $H_2$ plus $CO_2$, and seven major peaks in enrichments with lactate or ethanol. These results are in good agreement with microscopic observations of cell morphology of the corresponding enrichments.

**Immediate future work**

Continue characterizing the physiology of syntrophic association between each of two SRB strains and the methanogen:

1. Stability of association in several consecutive passages
2. Determine stoichiometry of substrates and metabolites (lactate consumption, and hydrogen, acetate, methane production)
3. Quantify cell numbers and biomass of methanogen and sulfate reducers.

Amplify and sequence 16S-23S intergenic spacer region and three other genes of *Desulfovibrio vulgaris* isolated from Lake Depue.

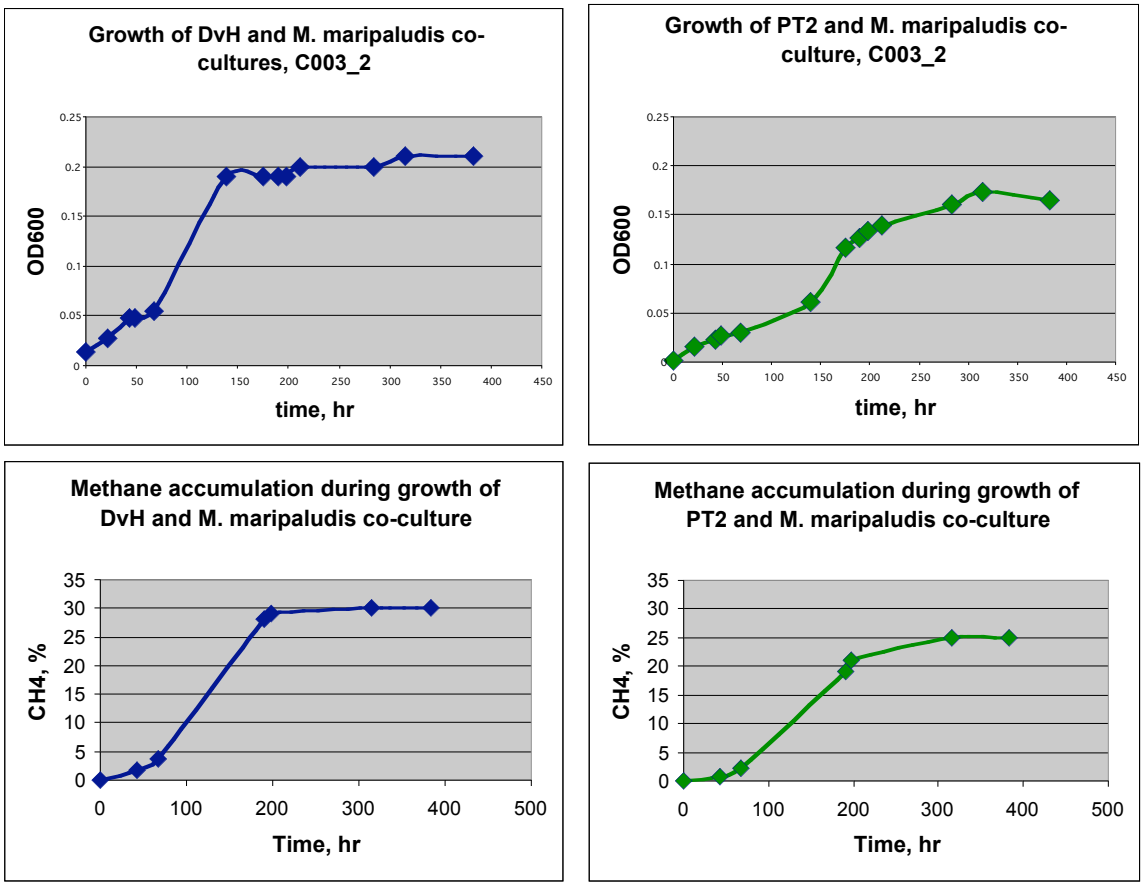
Continue isolation and characterization of sulfate reducing bacteria from FRC enrichments.

Amplify and sequence 16S rRNA and *dsrAB* genes from FRC isolates.

To study biofilm formation rates the new fixed bed reactor system consisting of one main (big) reactor and three smaller reactors will be inoculated again with a pure culture *D. vulgaris* Hildenborough. Small reactors will be opened (one reactor in a week) and biomass will be analyzed.

The FairMenTec Bioreactor inoculated with *D. vulgaris* Hildenborough will be tested in chemostat mode.

**Figure 1**.




***Oak Ridge National Laboratory***

Nothing new this month.

***Diversa***

**Progress:**

➢ Carl and Denise visited LBNL to perform large and small insert DNA extractions from the following $^{238}U$ contaminated soil samples:

Area 1: FB060-01-00, 0-25 inches, (2-20-03)
Area 2: FB052-01-00, near DP06, 20-25 feet, top, (2-18-03), ~200 ppm $^{238}U$.
Area 3: FB056-01-34, near FW009, (2-19-03), ~25 ppm $^{238}U$.

➢ Small and large insert libraries have been constructed for all previous GTL soil samples, and are ready to be screened.

**Issues:**

- Screening of libraries is dependent on selection of a suitable target.

**Actions:**

- Newly extracted DNA from samples collected at Areas 1, 2 and 3 are currently being amplified and will be used for library construction.

## III. Functional Genomics Core

# RESEARCH INTEGRATION OF THE CORE

## $O_2$ stress of *D. vulgaris* batch culture

### EXPERIMENT #1

Experiment #1 was carried out as described in the protocol on Monday the 15th of Sept. During the course of the experiment, several issues with biomass production and experimental set-up were identified and are described below. Solutions to remedy these problems are proposed along with protocols for experiments #3 and #4

1) The six batch cultures (triplicates of air and $N_2$ sparged cultures) were inoculated on Sunday the 14th at 9 am. The cultures reached 0.3 ODs at 1:30 pm.
2) Sampling of the cultures began at 1:30 and lasted about 2 h. All samples were taken from one batch culture before moving to the next. With the exception of Joe Zhou's samples, which were frozen in liquid nitrogen as they were sampled in the following order C3, C2, C1, V3, V2, V1), everyone (Dominique, Sharon, Swap, Aindrila, Vince and Lianhong) processed their respective samples all at once around 3:30 pm. Between 1:30 and 3:30, the individual samples were maintained in an anaerobic serum bottle at room temperature with the exception of Swap and Aindrila, which opted to cool their samples on ice first. In addition, all samples were briefly exposed to air during their transfer to falcon tubes for centrifugation and freezing. Note that we chose not to sample any biomass for Martin, as the cell counts were too low to give him the 1 to 2 g wet weight biomass that he requested. We opted to run a separate experiment for the proteomics samples.
3) The air and $N_2$ sparging was initiated once all the samples were removed from the batch cultures (at 3:30 pm). However, the reduction of the liquid in the culture bottles caused a loss in the gas stream though the medium. This was a result of the liquid level dropping below the porous tubing, causing the gas to flow out into the headspace (i.e. the path of least resistance). To overcome this problem, the three replicate cultures for each treatment were pooled into one bottle to raise the liquid level and restore gas flow through the liquid. Pooling of the cultures was done in the anaerobic chamber.
4) The two remaining pooled cultures were sparged for 5 h, at which time samples were taken and processed as described above.
5) The following control tests were done to ensure the quality of the biomass distributed to the various groups.

- Spread plate dilutions of the cultures on LS4D media (anoxic incubation) and TSA (oxic incubation) for verification of isolate purity via colony morphology. Results: no growth on aerobic incubations and typical colony morphologies on anaerobic incubations.
- AODC assay to determine direct counts. Results: typical cell morphology.
- DC protein assay to determine total biomass changes, and for isolate and biomass determinations

Table 1. OD of replicate batch cultures at T0 and T1

| **Time point** | **C-1** | **C-2** | **C-3** | **V-1** | **V-2** | **V-3** |
|---|---|---|---|---|---|---|
| **T0** | 0.3484 | 0.3334 | 0.3916 | 0.2496 | 0.3039 | 0.3133 |
| **T1 (5hrs)** | 0.5229 (filtered sample blank or 0.2900 (blank with uninoc. medium) | | | 0.7204 (filtered sample for blank or 0.4978 (uninoc medium blank) | | |

Table 2. Sample volumes taken per batch culture at time 0 =200ml

| | **T0**<br>**Cell Density: $5x10^7$** | **T1**<br>**Cell density: $10^8$** |
|---|---|---|
| Dominique/Sharon | 6x 15 ml | 2x 15 ml |
| Alyssa/Aindrila | 6x 50 ml | 2x 50 ml |
| Lianhong | 6x 10 ml | 2x 10 ml |
| Vince/Tung | 2x 50 ml | 2x 50 ml |
| Joe | 6x 50 ml | 2x 50 ml |
| Martin | None | None |
| Anup/Swap | 6x 50 ml | 2x 50 ml |

**Problems and solutions with experiment #1**

Problem #1: Sampling of the cultures takes too long.

The long sampling time was due to the fact that we were pulling samples from each anaerobic batch culture into anaerobic 150 mL serum bottles using a 60 mL syringe and aseptic technique. The long sampling period is very significant since a) Joe's first and last samples were taken ~2 h apart and b) everyone else's samples sat at room temperate in an anaerobic serum bottle for ~2hrs.

Solution: a) We can sacrifice smaller culture bottles at each time point. This would bypass the need to keep the cultures sterile and anaerobic while sampling and speed things up or b) set up a backpressure system with a valve on the flasks for rapid sampling of the cultures.

Problem #2: Biomass is too low.

As suspected from the growth curves, the biomass density is very low at T0 (OD ~0. 3 and $<10^8$ cells/mL). This creates a problem for the analysis of small molecules and proteome as 10 to 100 times more cell mass will be required.

Solution: Using batch cultures, we cannot move to higher cell density, as this will take us into stationary phase. Therefore, we have two possibilities 1) Use larger volumes or 2) move to chemostat cultures that would run at higher cell densities. We believe that 1) is the short-term answer but we will have to eventually move to chemostats.

Problem #3: Samples processing is not standardized

For this experiment it became clear that handling of the samples was not standardized. In particular, exposure time to air ($O_2$ stress) and cooling of the samples on ice (cold shock) were a concern. If measurements from these samples are to be compared, we need to have identical or at the very least similar sample handling procedures and we need to eliminate or at least reduce effects from any other stress other than the one we are trying to investigate.

Solution: a) Eliminate sample cooling on ice prior to processing.

- For proteomics samples, wash cells with room temperature degassed PBS once. This should reduce exposure to air, cold shock and sampling time.
- Process the metabolomics extraction and freeze the transcriptomics and proteomics samples at the exact same time.
- For other samples PLFA, QA/QC processing was not a problem.

Other issues: At this time, in order to shift to bigger culture volumes, we may need to sacrifice treatment replication within one experiment. In doing this, we would have to replicate each experiment three times. In a sense, this is a better experimental replication. Of course in a perfect world, replicates within and between experiments would be better. The other possibility is to run separate experiments for each measurement (i.e. one set of biomass for proteomics, one for transcriptomics etc….). This would create potential problems with comparing data sets from different experiments.

Following are two proposed experiments to incorporate some of the solutions for the problems we encountered during experiment #1. In addition, the experiments are designed to generate preliminary data of mRNA and protein profiles and do not attempt to get a complete data set for all measurements.

**Experiment #2** was setup and run on September 18th, to supply Martin alone his biomass. Unfortunately two of the replicates had been contaminated with air sometime during the night, since they were pink that morning due to the reazurin color change. Contamination was negative so we suspect it was caused by a tubing leak.

## EXPERIMENT #3

This experiment is designed to generate samples for Martin's proteomics analysis from two time points of a $N_2$ and air sparged grown batch culture, this experiment was successfully run on September 23rd.

The protocol for this experiment was as follows:

**Inoculum culture:**

- Freezer stock cells are transferred into fresh filter-sterilized lactate-sulfate defined anaerobic medium (LS4D) at a ratio of 1 ml inoculum/100 ml LS4D. Cells are grown for 48 h to log phase (OD of ~0.3; cells/ml). Note: freezer stock consists of log phase cells grown in BAARs and stored at –80°C.

**Biomass production culture:**

- Inoculate 720 mL of reduced lactate-sulfate (LS4D) anaerobic medium with 80 ml (10% v/v) of *D. vulgaris* from the log phase inoculum culture. Grow cultures in 1 liter glass bottles placed in 30°C water bath until an OD of 0.3 is reached (~$10^8$ cells/ml).
- Once the 800 mL cultures reach the desired OD, harvest 400 mL of cells for the time 0 (T0) sample (see protocols below for harvesting cells).
- After T0 cells are harvested, sparge one of the cultures with sterile humidified $N_2$ and the other culture with sterile humidified air.
- Grow the cultures at 30°C for one doubling time (5 h). Harvest the remaining amount of cells for the T1 sample.

**QA/QC for purity, biomass, and physiological state:**

As before

**Cell harvesting of *D. vulgaris* cultures for proteomics**

1. Weigh the cell-collecting tubes.
2. Harvest the cells by centrifugation at 3000 g for 10 min at room temperature.
3. Wash cells once with room temperature PBS buffer and harvest the cells again by centrifugation at 3000 g for 10 min at room temperature.
4. Remove the supernatant as much as possible to leave only cell pellet (remaining supernatant will affect protein extraction yield later). Weigh the tube again to calculate weight of cell pellets.
5. Quick freeze the sample in liquid nitrogen
6. Store them at -80°C and shipped on dry ice overnight

Sample codes:

| | |
|---|---|
| E3T0C-1 | E3T1C-1 |
| E3T0V-1 | E3T1V-1 |

## EXPERIMENT #4

This experiment is designed to generate time course samples for Joe's mRNA analysis from a $N_2$ and air sparged grown batch culture. Unfortunately the cultures were not processed properly and the biomass could not be used. This experiment is replicated as Experiment #7 (Oct progress report)

# PROGRESS OF EACH FACILITY

**Transcriptomics (ORNL)**

**Progress since last report**

*Shewanella*

- Manuscript describing the heat shock stimulon in *S. oneidensis* is close to being completed and will be submitted to the *Journal of Bacteriology*. The GTL Computational Core Group has completed an analysis of the upstream regions of HS-induced genes and has predicted the sigma-32 binding site in *Shewanella*. This site resembles the *E. coli* consensus sequence for sigma-32. This work is included in the

paper and demonstrates the collaborative effort between the different GTL functional groups.

- A manuscript describing the molecular response of *Shewanella* to low and high pH is currently in progress. We anticipate having this manuscript completed by the end of this year.
- We are continuing with microarray studies of the response to *Shewanella* to the heavy metal strontium. A time-series experiment is planned to examine the temporal gene response to 180 mM strontium. Preliminary growth studies showed no significant difference in strontium tolerance between wild type and a *fur* deletion mutant.
- Additional microarray hybridization experiments have been completed for the oxidative stress project. Growth studies and array data suggest that there is a relationship between oxidative-stress protection and iron homeostasis in *Shewanella*. Preliminary growth studies using Bioscreen indicate that a *fur* deletion mutant shows increased sensitivity to oxidative stress relative to the WT.
- A draft of the salt stress manuscript has been completed.
- Posters on *Shewanella* heat shock response, salt shock response, and metal tolerance were recently presented at the 11th International Conference on Microbial Genomes, held in Durham, NC.

Desulfovibrio

- *Desulfovibrio* cell samples were received from Terry Hazen's group.
- Cultivation conditions for *Desulfovibrio* have been explored here at ORNL. Sodium sulfide was found to be an appropriate reductant and enabled high growth yields for *Desulfovibrio*.
- Microarray experiments are planned to examine salt shock in *Desulfovibrio*.

Geobacter

- Oligonucleotide probes have been designed for approximately 4,200 genes in *Geobacter*. Probes need to be designed for the remaining 400-500 genes. Zhili He is working with Eric Alm on getting probes for the remaining genes.

## Proteomics (Diversa)

Objective

- Proteomics analysis of *D. vulgaris* stress-response upon exposure to air

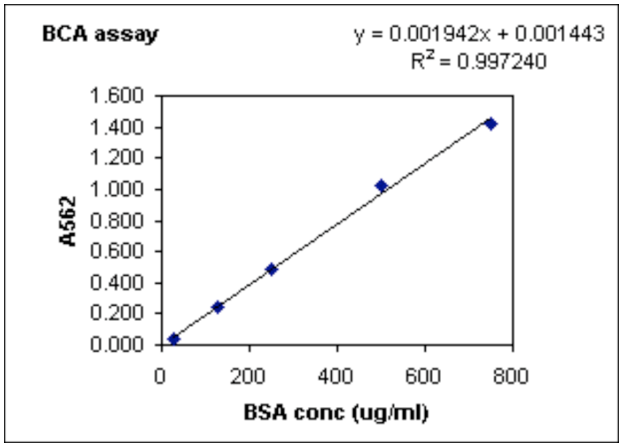
Progress since last report

***Point 1 and 2 performed by LBNL***

1. *Desulfovibrio vulgaris* cell pellets from experiment 3 with $O_2$ stress were received. *D. vulgaris* cells were cultured at 30C to reach OD600~0.3 ($10^8$ cells/ml) in two separate sets (C and V). Cells were harvested as **E3T0C1** and **E3T0V1** (should be the same). The cultures were sparged with sterile humidified $N_2$ (for C) or with sterile humidified air (for V). Grow the cultures for one doubling time (5h) and harvest the cells as **E3T1C1** and **E3T1V1**.
2. Cells were collected by centrifugation at 3,000 g for 10 min at room temperature and washed once with PBS buffer. Cells were frozen in liquid nitrogen and stored at –80C and shipped on dry ice overnight.

3. Thaw the cells and take out extra liquid. Measure the cell weight:
4. **E3T0C1: 15.36-14.78=0.58g**
   **E3T0V1: 15.25-14.81=0.44g**
   **E3T1C1: 15.33-14.84=0.49g**
   **E3T1V1: 15.33-14.75=0.58 g**
5. Suspend cells with 1.0 ml of 1% RapiGest in 50 mM Tris-HCl, pH 8.0. Transfer the fresh 1.5 ml tubes. Vortex and incubate at room temperature for 60 min to extract the proteins.
6. Centrifuge at 16,000 g for 10 min. Save the supernatants about 1.0 ml as E3T0C1-Rg, E3T0V1-Rg, E3T1C1-Rg, E3T1V1-Rg.
7. Measure protein concentration.

| BSA (ug/ml) | A562 |
|---|---|
| 25 | 0.033 |
| 125 | 0.247 |
| 250 | 0.484 |
| 500 | 1.021 |
| 750 | 1.427 |




**Sample analysis**

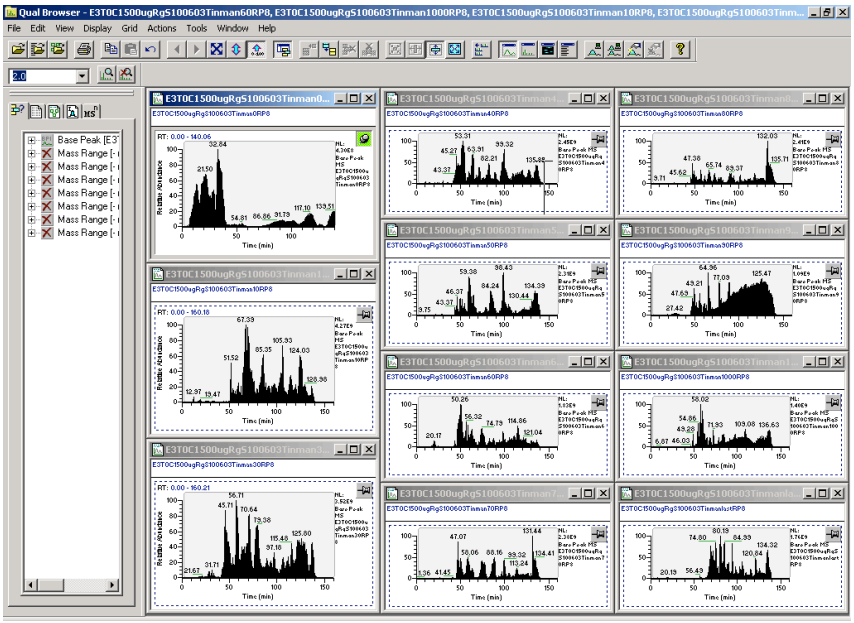
| Digested Sample | Protein Extract (mg) | Conc (ug/ul) | Digested Amount (ug) | Vol (ul) | Peptide samples |
|---|---|---|---|---|---|
| E3T0C1(X)a | 14 from 580 | 2.44 | 1000 | 410 | Rg whole cell from D vulgaris E3T0C1 |
| E3T0C1(X)b | | | | | Pellets after acid treated E3T0C1(X) |
| E3T0V1(X)a | 15 from 440 | 2.63 | 1000 | 380 | Rg whole cell from D vulgaris E3T0V1 |
| E3T0V1(X)b | | | | | Pellets after acid treated E3T0V1(X) |
| E3T1C1(X)a | 12 from 490 | 2.13 | 1000 | 470 | Rg whole cell from D vulgaris E3T1C1 |
| E3T1C1(X)b | | | | | Pellets after acid treated E3T1C1(X) |
| E3T1V1(X)a | 7 from 580 | 1.39 | 1000 | 720 | Rg whole cell from D vulgaris E3T1V1 |
| E3T1V1(X)b | | | | | Pellets after acid treated E3T1V1(X) |

3D LC MS/MS of following 8 samples

**Currently running**

| Digested Sample | Date loading | Sample name | Analysis time |
|---|---|---|---|
| E3T0C1(X)a | 10/5/2003-10/6/2003 | E3T0C1500ugRgS100603Tinman | 5.5 |
| E3T0C1(X)b | 10/6/2003-10/7/2003 | E3T0C1halfRgS100603Tiger | 3.5 |

Part of chromatograms of 3D LC MS/MS of E3T0C1(X)a



Future work

- Complete 3D LC MS/MS analysis of all 8 samples
- Data searching and analysis

## Protein complexes (Sandia)

Progress since last report

**Proteomic Analysis through DIGE in *D. vulgaris***

A. Heat shock response

B. Oxygen stress (Terry's lab)

Samples for both conditions were generated and whole cell protein quantified. We are awaiting 2D gel blots which will be then analyzed through MALDI.

**Protein Complex Analysis in *D. vulgaris* - Heat Shock response**

Based on the preliminary DIGE analysis of heat shock response in *D. vulgaris* HSP70 (ORF00281) was identified being involved in the stress condition (> 2 fold expression). Antibodies to the *E. coli* homolog (63% sequence identity) to ORF00281 were obtained from Upstate Biotech, MA. Western Analysis using this antibody on the extract from heat shock conditions showed enhanced production of ORF00281 (Hsp70).The Anit-HSP70 antibody was

then used to study bait-prey interactions in whole cell protein extracts from the heat shock condition using the Co-Immunoprecipitation kit (Invitrogen) for immobilization. "Pulled" down proteins were run on SDS-PAGE gels and silver stained - around 7 bands were observed - possibly interacting proteins with HSP70 under stress conditions. These bands were gel extracted and we are currently in the process of analyzing the extracted bands using MALDI.

**Mass spectrometry**

MALDI TOF instrument (DE PRO Voyager from Applied Biosystems) has been delivered, installed, and tested. In-gel digests of DIGE samples will be run as soon as the DIGE analysis is finished.

We have installed Mascot, the search engine used to analyze peptide mass fingerprint (PMF) data, on site and loaded the *D. vulgaris* database. We are currently optimizing parameters for batch search submission using Mascot Demon.

Future work

1. Identify other target proteins under stress conditions for bait-prey interaction studies.
2. Correlate DIGE and ICAT analyses for whole cell protein extracts. ICAT experiments will be done in Keasling lab.

## Metabolomics and Proteomics (UCB, LBNL)

Objectives

- Develop metabolite analysis using the newly acquired CE-MS
- Metabolite analysis of air-stressed *D. vulgaris* samples (Experiment#1)
- Construct and test vectors for directed gene knockout strategies in *D. vulgaris*.
- ICAT analysis of air-stressed *D. vulgaris* samples (Experiment#1)

Progress since last report

**Metabolomics**

The following samples were collected from experiment #1

**E1T0C-1: 10ml aliquots X 6 (extracted using the TCA-oil overlay method)**
**E1T0V-1: same**
**E1T1C-1: same**
**E1T1V-1: same**

Analysis of a few samples on the LC-MS showed that the metabolite concentrations were too low to analyze. This is not surprising as the cells had only reached an OD of $>10^8$ at sampling time. Methods are being developed to increase the sensitivity of these assays.

Method development:

- CE-MS separation of pyruvate, isocitrate, malate fumarate, succinate, 2-oxo-glutarate, asconitate, 3-phosphateglycerate and fructose 6-phosphate was accomplished (**Figure 1**)
- CE-MS separation of nucleotides was accomplished (**Figure 2**)
- CE-MS separation of amino acids, bases and nucleosides was accomplished. (**Figure 3**)

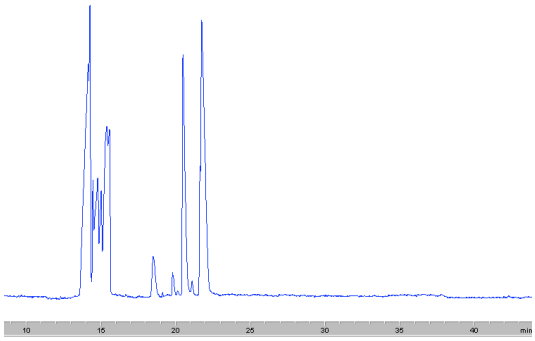


**Figure 1**. Separation of the metabolites in TCA and glycolysis pathways by CE-MS. A SMILE(+) column (50 μm × 120 cm, from Nacalai Tesque, Japan) was used. The signals were monitored by an Agilent single quadrapole mass detector in SIM scan mode (Single ion monitoring). The voltage is -30 kV. The electrolyte is 50 mM $NH_4OAc$, pH=9.0, and the sheath liquid is 5 mM $NH_4OAc$ in 50% MeOH-$H_2O$. ESI-MS was conducted in the negative ion mode and the capillary voltage was set at 3500 V. A mixture of 14 compounds was partially separated by CE and each compound can be monitored by the mass detector. The mixture contains the following metabolites: 2-oxoglutarate, isocitrate, cis-aconitate, citrate, oxaloacetate, malate, fumarate, succinate, glucose 6-phosphate, fructose 6-phosphate, 2-phosphoglycerate, phosphoenolpyruvate, pyruvate and ribose phosphate.

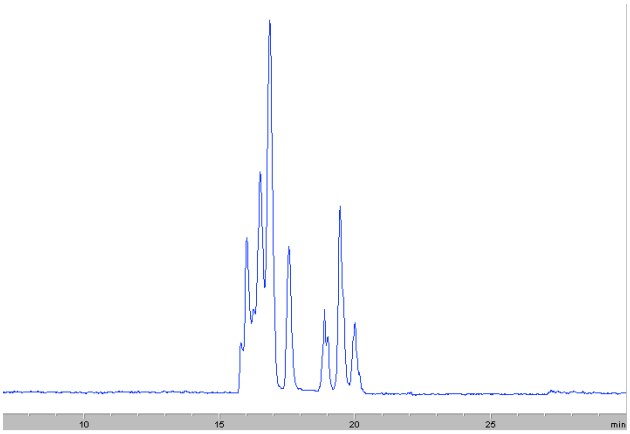
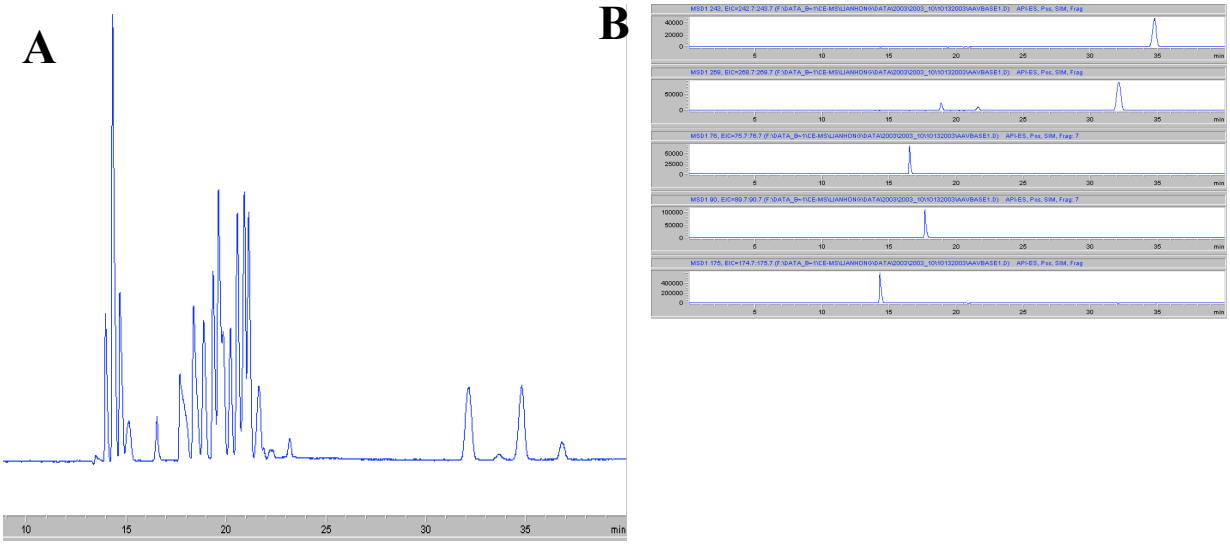


**Figure 2**. Separation of 19 nucleotides by CE-MS. A SMILE(+) column (50 μm × 120 cm, from ) was used. The signals were monitored by an Agilent single quadrapole mass detector in SIM scan mode (Single ion monitoring). The voltage is -30 KV. The electrolyte is 50 mM $NH_4OAc$, pH=9.0, and the sheath liquid is 5 mM $NH_4OAc$ in 50% MeOH-$H_2O$. ESI-MS was conducted in the negative ion mode and the capillary voltage was set at 3500 V. A mixture of 19 compounds was partially separated by CE and each compound can be monitored by the mass detector. The mixture contains the following metabolites: ATP, ADP, AMP, dATP, CTP, CDP, CMP, dCTP, GTP, GDP, GMP, dGTP, TTP, TDP, TMP, UTP, UDP, UMP and dUTP.



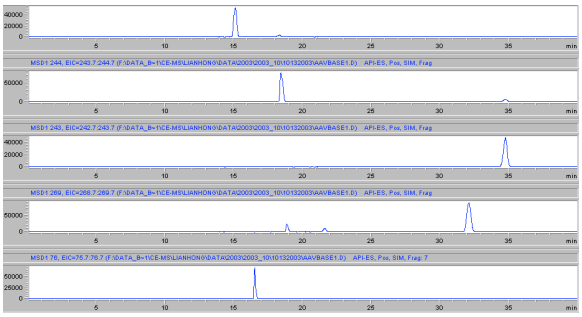
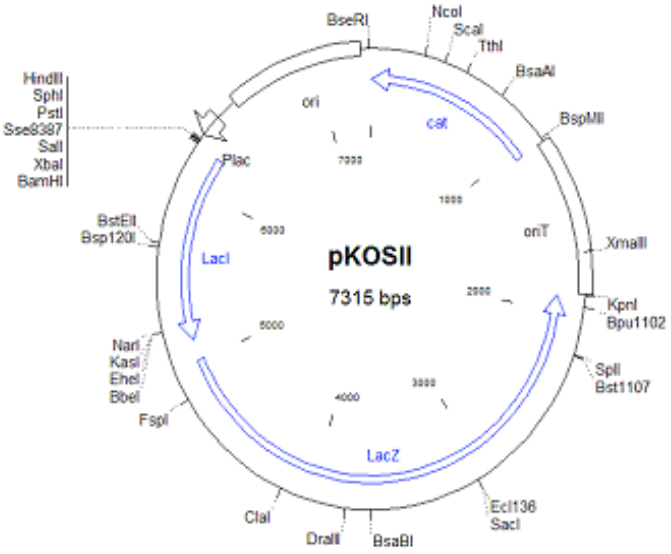


**Figure 3**. Separation of bases and amino acids by CE-MS **(A)**. A fused silica column (50 □m □ 120 cm) was employed. The signals were monitored by an Agilent single quadrapole mass detector in SIM scan mode (Single ion monitoring). The electrolyte is 1 M formic acid. The voltage is 30 KV. The sheath liquid is 5 mM $NH_4OAc$ in 50% MeOH-$H_2O$. ESI-MS was conducted in the positive ion mode and the capillary voltage was set at 4000 V. The mixture of the compounds was partially separated by CE and each compound can be monitored by the mass detector (B). The mixture contains the following metabolites: uridine, cytidine, guanosine, thymidine, adenine, inosine, xanthine, guanine, hypoxanthine, cytosine, adenosine, uracil, 2□-deoxycytidine, 2□deoxyadenosine, 2□deoxyuridine, 2□deoxyguanosine, xanthosine, glycine, alanine, valine, leucine, isoleucine, methionine, proline, cysteine, phenylalanine, tyrosine, tryptophan, arginine, lysine, histidine, aspartate, glutamate, serine, threonine, asparagines and glutamine.

**Gene knockout strategy II**

The final vector to conduct the single cross-over knockout mutagenesis has been successfully created. The vector, named pKOSII (Knock Out Strategy II) contains a chloramphenicol resistance cassette, an origin for maintenance in *E.coli*, an *oriT* for mobilization into other bacteria using *E.coli* mobilizer strain S17-1 and $P_{lac}$ driven full *lacZ lacI* genes that require both IPTG and X-gal for blue color.




To experimentally test the absence of galactosidase activity, WT *D. vulgaris* was plated on LS4D-Xgal medium and incubated until confluent growth was observed. Plates were taken out of the anaerobic chamber and further incubated at 4°C to allow color development. No Xgal activity was observed.

**ICAT Proteomics with *D. vulgaris***

Experiment 1: Samples collected on 9/12/03 from the Hazen Lab.

Samples:-

**E1T0C-1: 50ml aliquots X 6 (3 each for Keasling and Singh Labs)**
**E1T0V-1: same**
**E1T1C-1: same**
**E1T1V-1: same**
E1T1C-1 and E1T1V-1 are the control and $O_2$ stressed *D.vulgaris* samples at T=5hours. These were thawed and used as follows: -
The 6 aliquots for E1T1C-1 and E1T1V-1 were pooled and suspended in 6ml Tris-HCL pH 7.0
Lysed 2ml suspension via sonication
Lysed 2ml suspension via French press

Concentration of proteins in lysates:

| Sample | ug/ul |
|---|---|
| E1T1C-1 Sonicated | 11.9925 |
| E1T1C-1 French Press | 12.4925 |
| E1T1V-1 Sonicated | 8.2175 |
| E1T1V-1 French Press | 4.8175 |

The French Press lysates containing 100ug protein from E1T1C-1 and E1T1V-1 were used in an ICAT experiment as per manufacturer's instruction (Applied Biosystems). The ICAT samples are being currently analyzed using LCMSMS (QSTAR).

Future work

- Optimization CE-MS methods
- Quantization of compounds metabolites on CE-MS
- Development of methods for separation of more metabolites
- Application of these methods on *E. coli* cell extracts.
- To test *E.coli- D. vulgaris* conjugation protocols
- To conjugate pKOSII into another strain of *E.coli* to test its oriT functionality
- To conjugate pHRP308 (or another LacZ expressing plasmid) into *D. vulgaris* and test various fluorescent substrate alternatives to X-gal
- To clone 750bp internal regions of HSK(s) into the MCS region of pKOSII via hybridized or ligation.

## IV. Computational Core

***LBNL-Arkin***
In September, we completed a major update to our comparative genomics database, including all 127 finished genomes as well as several draft sequences. This includes "all vs. all" BLAST comparisons, domain identification, preliminary annotation of transposon and phage-related genes, operon and regulon predictions, and determination of putative orthology relationships

among the genes. In addition, this month we created gene models for a recent Geobacter metallireducens assembly, and assisted ORNL in picking a complete set of oligos for microarray design against this genome.

On the VIMSS comparative genomics wesite, we added a tool to allow download of a complete set of predicted metabolites for selected genomes to assist Vince Martin in his metabolomic experiments. In addition, the regulon browser on the protein pages now allows browsing of gene associations inferred from microarray data in addition to comparative genomic data when available.

We moved to new office space in September, and have purchased and installed new networking and file-sharing hardware. We hope to acquire additional hardware to facilitate data backup and transfer from our new location in the near future. In September, we hired and trained a new programmer, Morgan Price, who will work on implementing the comparative genomics data entry pipeline as well as improving operon predictions and predicting cis-regulatory elements. Morgan has already made significant progress in untrained operon prediction in E. coli, and we hope to make untrained predictions using our full scoring function for all sequenced bacterial genomes by next month.

This month, we produced an initial draft of operon-based quality assessment of individual microarray data sets for Zhou's group at ORNL. In addition, we have compiled figures comparing orthologous gene expression under similar conditions for E. coli and S. oneidensis for heat shock and low pH treatments. Using this comparative genomic and gene expression information, we made a predicted weight matrix for the sigma32 binding site in S. oneidensis.

Finally, we have begun planning for a (March/April) annotation jamboree at the Joint Genome Institute. In response to our request, JGI has moved our target organisms to the highest priority in their microbial finishing pipeline, and we have begun to explore how best to implement our own tools in the JGI annotation system.

***LBNL – Olken***

## Visual Graph Query Language

Work continued on design of a visual graph query language user interface. We have been searching for toolkit for building the visual graph query interface. JGraph and IsaViz appear to be the two most promising candidates and we are leaning toward use of IsaViz, which is already an RDF editor and also has facilities for graph style sheets, which should facilitate graph visualization.

## Graph Data Model

We continue development of the graph data model, working on details of RDF representation of common features of biopathways graphs. With the first release of the Biopax schemas for biopathway data interchange in OWL/RDF we have begun to address use of this data interchange standard.

## Navigational API

Work continues on the design of the navigational API. In Java. This API will be used by applications (or GUIs) to navigate amidst the graph results. Portions of this API are being tested in the browser development (see below).

## Constructing RDF Query Graphs

We have decided to encode the graph queries in RDF. We are presently working on detailed specification of the query representation in RDF. We have decided that we use an existing RDF graph constructor library for this purpose. Initial investigation suggests the use of the Jena library from HP, but we are also looking at other libraries.

## Browser

Development of the database schema and instance browser by Kevin Keck continued. This browser permits the user to view a relational database schema as a graph. We anticipate that it will be used in conjunction with the graph query tool to assist in query formulation. It is also a testing group for the navigational API. We are successfully accessing the schema from BIODB database via JDBC. Many of the browser functions are working. We are continuing development of presentation and formatting of links.

## Writing

A near final draft version of the workshop report on the Workshop on Data Management for the Life Sciences (Feb. 2-3, 2003) was sent to the writing committee. We hope to have the final report out at the end of October.

## Staffing

Vija Natarajan (from Scientific Data Management Group) will join the project in mid-October to work on the visual graph query language interface. She is being supported from Synechoccocus GTL funds.

## Plans for October

1) Continue development of RDF/web based schema browser prototype. (KK)
2) Continue development of RDF/web based relational DB instance browser. (KK)
3) Complete biopathways chapter for DOE Computational Biology Primer. (FO)
4) Complete report on NLM Workshop on Data Management.(FO)
5) Complete Navigational API specification. (KK)
6) Code Navigational API. (KK)
7) Code DOT file format (from ATT Graphviz) to RDF converter. (KK)
8) Select of graph toolkit will be used for VGQL. (VN, KK, FO)
9) Selection of RDF graph constructor toolkit for VGQL. (VN, KK, FO)
10) Specification of graph query language RDF encoding. (FO, KK)

11) Commence algorithm design for path queries which satisfy a regular expression. (FO)
12) Begin efforts to develop a common biopathway data interchange format with other members of Arkin Lab, the Ozsoyoglu's at Case Western Reserve Univ, Peter Mork (Univ. of Washington) and biopax.org. (FO,KK)
13) Attend Semantic Web Meeting. (KK)
14) Host and attend Synechoccocus GTL meeting at LBNL on Oct. 21-22. (FO)

# V. Project Management

### FY04 Budget

The FY04 budget is under development. The DOE has not given any guidance regarding the status of the budget. A Continuing Resolution budget is anticipated until the federal budget is approved.